

RECEIVED

DEC 11 2000

1652

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/463,705A

DATE: 12/07/2000
 TIME: 12:39:57

TECH CENTER 1600/2900

Input Set : A:\Cpg.pto
 Output Set: N:\CRF3\12072000\I463705A.raw

3 <110> APPLICANT: Cortes, J
 4 Promentin, C
 5 Gaisser, S
 6 Leadlay, P
 7 Mendez, C
 8 Michel, J-M
 9 Raynal, M-C
 10 Salah-Bey, K
 11 Salas, J
 13 <120> TITLE OF INVENTION: BIOSYNTHESIS GENES & TRANSFER OF 6-DESOXYHEXOSES IN
 14 SACCHAROPOLYSPORA ERYTHRAEA AND IN STREPTOMYCES
 15 ANTIBIOTICUS AND THEIR USE
 17 <130> FILE REFERENCE: 146.1335
 19 <140> CURRENT APPLICATION NUMBER: 09/463705A
 20 <141> CURRENT FILING DATE: 2000-02-23
 22 <150> PRIOR APPLICATION NUMBER: PCT/FR98/01593
 23 <151> PRIOR FILING DATE: 1998-07-21
 25 <150> PRIOR APPLICATION NUMBER: 98/07411
 26 <151> PRIOR FILING DATE: 1988-06-12
 28 <150> PRIOR APPLICATION NUMBER: 97/09458
 29 <151> PRIOR FILING DATE: 1997-07-25
 31 <160> NUMBER OF SEQ ID NOS: 61
 33 <170> SOFTWARE: PatentIn Ver. 2.1
 35 <210> SEQ ID NO: 1
 36 <211> LENGTH: 3439
 37 <212> TYPE: DNA
 38 <213> ORGANISM: Saccharopolyspora erythraea
 40 <220> FEATURE:
 41 <221> NAME/KEY: CDS
 42 <222> LOCATION: Complement((48)..(1046))
 43 <223> OTHER INFORMATION: /function= "involved in the biosynthesis of
 44 mycarose" /gene= "eryBII"
 46 <220> FEATURE:
 47 <221> NAME/KEY: CDS
 48 <222> LOCATION: Complement((2322)..(3404))
 49 <223> OTHER INFORMATION: /function= "involved in the biosynthesis of
 50 desosamine" /gene= "eryCII"
 52 <400> SEQUENCE: 1
 53 gcttcacgct caccagccgt atcctttctc ggttcctctt gtgtcactg caaccaggt 60
 55 tccgycgccc cgccgcccga ggccaccgag gggaagatct cgtccagttc ggacagcgcc 120
 57 tgctcgtcca ggttcctcgc ggacgccttc agcgcggagt cgagctgctc ggggggttcgc 180
 59 gggccgatga cggcgccggc gatgccgggc cgggacagca cccatgcyag ccccaactcg 240
 61 gccgggtctt cgccgaggtt gcggcagaac ttctcgtagg cctcgtatgc cgggcgcagg 300
 63 gacggcaaca gcaactgcgc acggccctgc gccgacttca ccgcggtgcc cgcggccagc 360
 65 ttctccagcg ctccgctyag caggcccgcc tgcagcgccg accaggcgaa gacgcccagc 420
 67 ccgtaggcct gcggcgcygg cagcacttcc agctcggcgt gccggaccgc caggttgatc 480
 69 aggcactggg gggagaccat gccagggag tggcgccggg cggcgcttctc ctgcgcggcg 540

RAW SEQUENCE LISTING

DATE: 12/07/2000

PATENT APPLICATION: US/09/463,705A

TIME: 12:39:57

Input Set : A:\Cpg.pto

Output Set: N:\CRF3\12072000\I463705A.raw

RECEIVED

DEC 11 2000

TECH CENTER 1600/2900

```

71 gcgatgtgcc agcccgcgaa gttegcagag ccgacgtagg agaccttgcc gclggcgagc 600
73 aggctgtcca tggcctgcca caccctctcc cacygcgcgc accggtcgat gtggtgcac 660
75 tggtagacgt cgtatgtgtc gacgcccagc ctgcgcagcg atccctcgca ggaggcgatg 720
77 atgtgccgcg ccgacagccc gctgtcgttg acgcgctcgc tcatctcgcc gccgaccttg 780
79 gtgcgcagca cgggtgtctc gcgcgctccg ccgcctcgcg ccagccacct gccaccaagc 840
81 tctcgggtgt ggccttgtta gagccgcagc ccgtacatgt cggcggtgtc gaggcagttg 900
83 atgcgcgcgt ccggggcgtg gtccatcagg cgcacgcgct cgtcgtctcc gacgcgtccg 960
85 ctgaagtcca ccgtgcgcag ccagagcctg ctggtgagca gcgcggaacg ccgagagccg 1020
87 acgtgcgtcg cggcgctcgt ggtcatctgt gttctctctt tccgtcgccg agttccctcg 1080
89 agatgcgcag gacctcgccc ggtgacggct ccgcgagcat gtcgtcgcgc atccgcgcgc 1140
91 cgcgcgcgcg gtggggcggg tctgtcagga ccgccttcac cgaactcccg agctggtcgg 1200
93 ggttcagctc gggcacgggc agcgcgatcc ccgcgccgaa ttctgcgtg cgtgcgcgcg 1260
95 gcacgcgcgt gtcccagccg tcgggcagga tcacctgcgc cagcgcgtg atgcgcgcgc 1320
97 tgtgcgcagt ccggggtccg ccgtggtgca ccgtgcgcgc gcaggtcgcg agcagcgct 1380
99 gcctcgagac gaagccgacc gtgcggagct tglccgggat gttcgcgagc ccttctagct 1440
101 cctgcgcgtc gaaggtcgcg atgatctcgc cgtcgacgtc gccgacgca cccagcagct 1500
103 cctcgatgga gacctgccc atgctgttct cgcggtgga gatcccgagc gtgaggcaca 1560
105 cgcgcgcgcg ctgcggtcgt tctgtcagcc attccgcac cagcgacygc ccgttctagt 1620
107 cgacgtagcg catcccgacc gtcttcaggc cgtgtcgcag cctgatcgcg gccggggcgg 1680
109 ggtcgatcgt ccactgccc acgaccacct cctcgtcgaa ggcggggcgc cgtacttct 1740
111 ccagcgtcca ggtgagccac tcggcgagcg ggtcctcccg gtgctcctcc ggtggtcg 1800
113 gcagcagggc gaggaaagtt tgcgcgcgcc ggggtggtat gtcgggtccc cagcagacc 1860
115 gcgcgtgcgc cgttcgggtc accgcgcgcg cgtggtgcgc ggcgaagtg agcggtccc 1920
117 agatgaccag gtcgggcgc cacttccggc agaacgagac catgcttcg atgagcgtgt 1980
119 ccgggtctat cagggcgtag aaggtcgggg tgaacacggt ctgcatcgcc agcaggtgct 2040
121 ccaggttcaa ggtggtgggg tccgcgtcgc tgaagtccag gctccgagc tagtcgatga 2100
123 tgtcgtggcc cgcgtgggtc atgaagttca cgaagtcgac gtcggtgcgc accgggacgg 2160
125 cgttcagccc ggcgcgcgtg atgtctcgc tgaagcgcgc ggcgcgcacc acgcgcacct 2220
127 cgtgcgcgcg cgcgcggaac gcccatgcca gggggagcag gccgaagag tggctcttgc 2280
129 tggccatgga ggagaagacg acgcgcacgc cgtttacctc agagctcgac ggggcagcgc 2340
131 ttggttcccc gcaggacggg tgatcgcgcg cgcgcgacga ccgggcgcgt gggcgtgagt 2400
133 ccgggcagcg ccttgcgcgc ggcgcgcagt gcgcgcgtg cagcgcggtg gaccagctcc 2460
135 tccagcctgc cgggtggtgc gcgatgtgcc gacagcgcg cgtcgcgctc ggggcgggtc 2520
137 acgtcgaggg ggtcggtgct ggcgaagacc tccgggtcgc ggttggcgc cgcgacgacg 2580
139 accagacact cctgccttcc gccgatcagc tgcgcgcga gccgcacctc tgcggtgacc 2640
141 gtgcgcgcgt ccaggtgcaa tgcgggtgac aggcgcagca cctcggcgac ggttcgctgc 2700
143 gcggcgcgcg ggtcgtcggc gatcgttccg gccagcccg gttcgccga gacggccagg 2760
145 accgcgtcga ccacggtgtt cgcggtcatc tgcgcccccg cgaacagggc gcgagtgcg 2820
147 gggtcggcgg gcagtgccgc gacctgtgtc tgggtcacgc cgaagctgtg cgggttgagc 2880
149 tggcgctcca ggtgagcgcg ggcgtccac gcggcgccgc gcagcactcc ggtgcgcgcg 2940
151 agcagcgcgg tcatgcccgt caccggtacc tgcagggcga agtcgcgac caggtccagc 3000
153 cgcgcgcgcg cgcgggggag cagaccggcg aagctctccg ccagttcccc gacgtcgggg 3060
155 acctgcctt cccaggacgc ggcgtgcagc tcccgaaacg gctgggccc ctcggcggt 3120
157 ggcgcgcgcg cggcccgcat ccattccggt gtgcgtcccg tggcgcggtg gaacgcgggg 3180
159 tctcgagcga cctgcgggcy ggtggcggtg tgcgcccca cccagctctc ggtgcggtg 3240
161 cgcgcgcac cgaactcgcg catcgagcgc taccgycgct gcgggtgctc gtcgtgtccg 3300
163 cacagcagca tgggttaagg gtgcgcgttg ctgcgtaac ccaagtcag gccgcggtc 3360
165 atctggagct gctgcgcag cccgcgcgca tgggtcgttg tcatgaattc cctccgcca 3420
167 gccaggcgct gatgtgccg

```

3439

RAW SEQUENCE LISTING

DATE: 12/07/2000

PATENT APPLICATION: US/09/463,705A

TIME: 12:39:57

Input Set : A:\Cpg.pto

Output Set: N:\CRF3\12072000\I463705A.raw

RECEIVED

DEC 11 2000

TECH CENTER 1600/2900

```

170 <210> SEQ ID NO: 2
171 <211> LENGTH: 333
172 <212> TYPE: PRT
173 <213> ORGANISM: Saccharopolyspora erythraea
175 <400> SEQUENCE: 2
176 Met Thr Thr Asp Ala Ala Thr His Val Arg Leu Gly Arg Ser Ala Leu
177 1 5 10 15
179 Leu Thr Ser Arg Leu Trp Leu Gly Thr Val Asn Phe Ser Gly Arg Val
180 20 25 30
182 Glu Asp Asp Asp Ala Leu Arg Leu Met Asp His Ala Arg Asp Arg Gly
183 35 40 45
185 Ile Asn Cys Leu Asp Thr Ala Asp Met Tyr Gly Trp Arg Leu Tyr Lys
186 50 55 60
188 Gly His Thr Glu Glu Leu Val Gly Arg Trp Leu Ala Gln Gly Gly Gly
189 65 70 75 80
191 Arg Arg Glu Asp Thr Val Leu Ala Thr Lys Val Gly Gly Glu Met Ser
192 85 90 95
194 Glu Arg Val Asn Asp Ser Gly Leu Ser Ala Arg His Ile Ile Ala Ser
195 100 105 110
197 Cys Glu Gly Ser Leu Arg Arg Leu Gly Val Asp His Ile Asp Val Tyr
198 115 120 125
200 Gln Met His His Ile Asp Arg Ser Ala Pro Trp Asp Glu Val Trp Gln
201 130 135 140
203 Ala Met Asp Ser Leu Val Ala Ser Gly Lys Val Ser Tyr Val Gly Ser
204 145 150 155 160
206 Ser Asn Phe Ala Gly Trp His Ile Ala Ala Gln Glu Asn Ala Ala
207 165 170 175
209 Arg Arg His Ser Leu Gly Met Val Ser His Gln Cys Leu Tyr Asn Leu
210 180 185 190
212 Ala Val Arg His Ala Glu Leu Glu Val Leu Pro Ala Ala Gln Ala Tyr
213 195 200 205
215 Gly Leu Gly Val Phe Ala Trp Ser Pro Leu His Gly Gly Leu Leu Ser
216 210 215 220
218 Gly Ala Leu Glu Lys Leu Ala Ala Gly Thr Ala Val Lys Ser Ala Gln
219 225 230 235 240
221 Gly Arg Ala Gln Val Leu Leu Pro Ser Leu Arg Pro Ala Ile Glu Ala
222 245 250 255
224 Tyr Glu Lys Phe Cys Arg Asn Leu Gly Glu Asp Pro Ala Glu Val Gly
225 260 265 270
227 Leu Ala Trp Val Leu Ser Arg Pro Gly Ile Ala Gly Ala Val Ile Gly
228 275 280 285
230 Pro Arg Thr Pro Glu Gln Leu Asp Ser Ala Leu Lys Ala Ser Ala Met
231 290 295 300
233 Thr Leu Asp Glu Gln Ala Leu Ser Glu Leu Asp Glu Ile Phe Pro Ala
234 305 310 315 320
236 Val Ala Ser Gly Gly Ala Ala Pro Glu Ala Trp Leu Gln
237 325 330
240 <210> SEQ ID NO: 3
241 <211> LENGTH: 361

```

RAW SEQUENCE LISTING

DATE: 12/07/2000

PATENT APPLICATION: US/09/463,705A

TIME: 12:39:57

Input Set : A:\Cpg.pto

Output Set: N:\CRF3\12072000\I463705A.raw

```

242 <212> TYPE: PRT
243 <213> ORGANISM: Saccharopolyspora erythraea
245 <400> SEQUENCE: 3
246 Met Thr Thr Thr Asp Arg Ala Gly Leu Gly Arg Gln Leu Gln Met Ile
247 1 5 10 15
249 Arg Gly Leu His Trp Gly Tyr Gly Ser Asn Gly Asp Pro Tyr Pro Met
250 20 25 30
252 Leu Leu Cys Gly His Asp Asp Asp Pro Gln Arg Arg Tyr Arg Ser Met
253 35 40 45
255 Arg Glu Ser Gly Val Arg Arg Ser Arg Thr Glu Thr Trp Val Val Ala
256 50 55 60
258 Asp His Ala Thr Ala Arg Gln Val Leu Asp Asp Pro Ala Phe Thr Arg
259 65 70 75 80
261 Ala Thr Gly Arg Thr Pro Glu Trp Met Arg Ala Ala Gly Ala Pro Pro
262 85 90 95
264 Ala Glu Trp Ala Gln Pro Phe Arg Asp Val His Ala Ala Ser Trp Glu
265 100 105 110
267 Gly Glu Val Pro Asp Val Gly Glu Leu Ala Glu Ser Phe Ala Gly Leu
268 115 120 125
270 Leu Pro Gly Ala Gly Ala Arg Leu Asp Leu Val Gly Asp Phe Ala Trp
271 130 135 140
273 Gln Val Pro Val Gln Gly Met Thr Ala Val Leu Gly Ala Ala Gly Val
274 145 150 155 160
276 Leu Arg Gly Ala Ala Trp Asp Ala Arg Val Ser Leu Asp Ala Gln Leu
277 165 170 175
279 Ser Pro Gln Gln Leu Ala Val Thr Glu Ala Ala Val Ala Ala Leu Pro
280 180 185 190
282 Ala Asp Pro Ala Leu Arg Ala Leu Phe Ala Gly Ala Glu Met Thr Ala
283 195 200 205
285 Asn Thr Val Val Asp Ala Val Leu Ala Val Ser Ala Glu Pro Gly Leu
286 210 215 220
288 Ala Glu Arg Ile Ala Asp Asp Pro Ala Ala Ala Gln Arg Thr Val Ala
289 225 230 235 240
291 Glu Val Leu Arg Leu His Pro Ala Leu His Leu Glu Arg Arg Thr Ala
292 245 250 255
294 Thr Ala Glu Val Arg Leu Gly Glu His Val Ile Gly Glu Gly Glu Glu
295 260 265 270
297 Val Val Val Val Val Ala Ala Ala Asn Arg Asp Pro Glu Val Phe Ala
298 275 280 285
300 Glu Pro Asp Arg Leu Asp Val Asp Arg Pro Asp Ala Asp Arg Ala Leu
301 290 295 300
303 Ser Ala His Arg Gly His Pro Gly Arg Leu Glu Glu Leu Val Thr Ala
304 305 310 315 320
306 Leu Ala Thr Ala Ala Leu Arg Ala Ala Ala Lys Ala Leu Pro Gly Leu
307 325 330 335
309 Thr Pro Ser Gly Pro Val Val Arg Arg Arg Ser Pro Val Leu Arg
310 340 345 350
312 Gly Thr Asn Arg Cys Pro Val Glu Leu
313 355 360

```

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/463,705A
 DATE: 12/07/2000
 TIME: 12:39:57

Input Set : A:\Cpg.pto
 Output Set: N:\CRF3\12072000\I463705A.raw

317 <210> SEQ ID NO: 4
 318 <211> LENGTH: 1266
 319 <212> TYPE: DNA
 320 <213> ORGANISM: Saccharopolyspora erythraea
 322 <220> FEATURE:
 323 <221> NAME/KEY: CDS
 324 <222> LOCATION: complement (4)..(1266)
 325 <223> OTHER INFORMATION: /function= "involved in the biosynthesis of
 326 mycarose" /gene= "eryCIII" /note= "SEQ ID No 1
 327 from 1046 to 2308"
 329 <400> SEQUENCE: 4
 330 tcatacgtggt tctctccttc ctgcggccag ttctctgcag atgcgcagca cctcggccgg 60
 332 tgacgggtcc gcgagcatgt cgtcgcgcgt ccgcgcgcgc ccggcgcggt gggccgggtc 120
 334 gtgcaggacc cgcttcaccg actcccgagg ctggtcgggg gtacgctcgg gcacggggcg 180
 336 cgcgaccccc gcccgaatt cctgcgtgag ctgcgcgcgc acgcgcggtg cccagccgct 240
 338 qqcaagatc acctgcggca cgcggtgat cgcgcggtg tgccagctcc cgggtccgcc 300
 340 gtgtgtgacc gtgcgcgcgc aggtcggcag cagcgcgtgc atcgggacga agccgaccgt 360
 342 gcggacggtt gccgggatgt tcgcgcgcgc ttctagctgc tgcgcgctga aggtcgcgat 420
 344 gatctcggcg tcgacgtgc cgcgcgcacc cgcagctcc tcgatggaga cctgcccgat 480
 346 gctgttctcg cggctggaga tcccgagcgt qagccacacg cggcggcgct cgggctcgtc 540
 348 gtgcagccat tccggcacca cgcgcgcgc gttgtagtcg acgtagcgca tcccgacggt 600
 350 cttcaggccg gtgtcgagcc tgatcgcggc cggggcgggg tcgatcgtcc actgcccgac 660
 352 gaccacctcc tcgtcgaagg ccgggcgcgc gtactctctc agcgtccagg tgagccactc 720
 354 ggcgagcggg tctctccggt gctctccggt ctggtcgggc agcagggcga ggaagtctc 780
 356 ccgcgcgcgc gtygtgatgt cgggtcccca cagcagccgc gcgtcgcgcg ttcgggtcac 840
 358 cgcgcgcgcg atgggcgcgc cgaaggtgag cggctccag atgaccagg cgggcgcgca 900
 360 cttccggcag aacgagacca tgccttcgat gagcgtgtcc gggctcatca gggcgtagaa 960
 362 ggtcgggggt agcacggtct gcattccag caggtgtctc caggtcaagg tggcgggggt 1020
 364 ccgctcgtcg aagtccaggc tccggacgta gtcgatgatg tcgtgycceg cgtgggtcat 1080
 366 gaagtcacag aggtcgacgt cgggtccgac cgggacggcg gtcagcccg cgcgggtgat 1140
 368 gtctcgtgtg agcgcggggg acgcgaccac gcggacctcg tgcgcccgcg cgcggaacgc 1200
 370 ccatgcgagg gggacgagg cgaagagggt gctcttctg gccatggagg agaagacgac 1260
 372 gcgcac 1266
 375 <210> SEQ ID NO: 5
 376 <211> LENGTH: 421
 377 <212> TYPE: PRT
 378 <213> ORGANISM: Saccharopolyspora erythraea
 380 <400> SEQUENCE: 5
 381 Met Arg Val Val Phe Ser Ser Met Ala Ser Lys Ser His Leu Phe Gly
 382 1 5 10 15
 384 Leu Val Pro Leu Ala Trp Ala Phe Arg Ala Ala Gly His Glu Val Arg
 385 20 25 30
 387 Val Val Ala Ser Pro Ala Leu Thr Glu Asp Ile Thr Ala Ala Gly Leu
 388 35 40 45
 390 Thr Ala Val Pro Val Gly Thr Asp Val Asp Leu Val Asp Phe Met Thr
 391 50 55 60
 393 His Ala Gly His Asp Ile Ile Asp Tyr Val Arg Ser Leu Asp Phe Ser
 394 65 70 75 80
 396 Glu Arg Asp Pro Ala Thr Leu Thr Trp Glu His Leu Leu Gly Met Gln

VERIFICATION SUMMARY

DATE: 12/07/2000

PATENT APPLICATION: US/09/463,705A

TIME: 12:39:58

Input Set : A:\Cpg.pto

Output Set: N:\CRF3\12072000\I463705A.raw

L:42 M:351 W: Sequence data Name/Key Feature Out-of-Range, SEQ ID#:1, CDS LOCATION: Complement((48)..(1046))
L:48 M:351 W: Sequence data Name/Key Feature Out-of-Range, SEQ ID#:1, CDS LOCATION: Complement((2322)..(3404))
L:324 M:351 W: Sequence data Name/Key Feature Out-of-Range, SEQ ID#:4, CDS LOCATION: complement (4)..(1266)
L:602 M:361 W: Invalid Split Codon, Sequence data for SEQ ID#: 6
L:603 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:6
L:607 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:6
L:944 M:361 W: Invalid Split Codon, Sequence data for SEQ ID#: 6
L:1068 M:361 W: Invalid Split Codon, Sequence data for SEQ ID#: 6

M Rao
1652

Serial Number: 09463, 705A

ENTERED

RECEIVED

DEC 11 2001

#15

TECH CENTER 1600/2900

- ☐ Changed a file from non-ASCII to ASCII
- ☐ Changed the margins in cases where the sequence text was "wrapped" down to the next line.
- ☐ Edited a format error in the Current Application Data section, specifically: _____
- ☐ Edited the Current Application Data section with the actual current number. The number inputted by the applicant was ☐ the prior application data; or ☐ other _____
- ☐ Added the mandatory heading and subheadings for "Current Application Data".
- ☐ Edited the "Number of Sequences" field. The applicant spelled out a number instead of using an integer.
- ☐ Changed the spelling of a mandatory field (the headings or subheadings), specifically: _____
- ☐ Corrected the SEQ ID NO when obviously incorrect. The sequence numbers that were edited were: _____
- ☐ Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited: _____
- ☐ Corrected subheading placement. All responses must be on the same line as each subheading. If the applicant placed a response below the subheading, this was moved to its appropriate place.
- ☐ Inserted colons after headings/subheadings. Headings edited included: _____
- ☐ Deleted extra, invalid, headings used by an applicant, specifically: _____
- ☐ Deleted: ☐ non-ASCII "garbage" at the beginning/end of files; ☐ secretary initials/filename at end of file; ☐ page numbers throughout text; ☐ other invalid text, such as _____
- ☐ Inserted mandatory headings, specifically: _____
- ☐ Corrected an obvious error in the response, specifically: _____
- ☐ Edited identifiers where upper case is used but lower case is required, or vice versa.
- ☐ Corrected an error in the Number of Sequences field, specifically: _____
- ☐ A "Hard Page Break" code was inserted by the applicant. All occurrences had to be deleted.
- ☐ Deleted *ending* stop codon in amino acid sequences and adjusted the "(A)Length:" field accordingly (error due to a PatentIn bug). Sequences corrected: _____
- ☒ Other: Corrected format error: inserted "hard returns"
at <211> feature for sequences 32 and 50.

*Examiner: The above corrections must be communicated to the applicant in the first Office Action. DO NOT send a copy of this form.

3/1/95